

From a Registry to a Registry Ecosystem

CommonsDB
and the future
of EU copyright
infrastructure

September 2026

Authors

Paul Keller Open Future

Sebastian Posth Liccium

commons DB



CommonsDB is an initiative funded by the European Commission to develop and test a prototype registry of Public Domain and openly licensed works. CommonsDB is led by [Open Future](#) in collaboration with [Liccium](#), [Europeana Foundation](#), [Wikimedia Sverige](#), and the [Institute for Information Law](#) (IViR).



This work has been co-funded by the European Union under the 2023 work programme on the financing of Pilot Projects and Preparatory Actions in the field of Communications Networks, Content and Technology. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.



This study is published under the terms of the [Creative Commons Attribution License](#).

Contents

1	Introduction	4
2	What CommonsDB has demonstrated	7
3	From single registries to an ecosystem	12
4	A federated architecture for copyright-information registries	16
	Two approaches to building a European registry infrastructure	17
	Records belong close to those who make them	18
	The federated model can deliver what registry users need	19
	Certification: what it takes for a record to count	20
	The envisaged architecture	21
	Costs and incentives	22
5	The EUIPO's role: certify, aggregate, anchor – and migrate	24
6	Conclusions	28
	Notes	30

1

Introduction

For most of its history, European copyright law has been defined by the absence of formal registration requirements. The Berne prohibition on formalities¹ made the absence of registries a structural feature of the system: protection arises automatically, and no public record exists of what is protected, by whom, or under what conditions. This has never meant an absence of records altogether: where rights are managed collectively or at scale, professional intermediaries – collective management organisations above all – maintain extensive documentation of works, rightsholders and mandates, connected through international networks. But these records are internal to the sectors that produce them, serve the administration of rights rather than public information, and cover only works that are actively managed. For as long as copyright governed a relatively slow-moving world of professional intermediaries, this arrangement held: the costs of the missing public record were absorbed by those intermediaries and their documentation systems.

However, in a world where automated systems make decisions relevant to copyright about billions of works (filtering uploads, assembling AI training corpora, resolving reuse permissions) the absence of reliable, machine-readable rights information has become an increasingly hard-to-ignore issue. As a result, European copyright policy has an unresolved infrastructure problem.

Member states now acknowledge this diagnosis. The [Copyright Infrastructure Task Force](#), a multi-country expert forum that addresses copyright data interoperability and trust, was formed specifically to tackle it². Most of the remedies under discussion share a common foundation: machine-readable rights registries.

Much of the current interest in registries is driven by the challenges raised by generative AI, and in particular by the text and data mining opt-out of Article 4(3) of the CDSM Directive: five years after the Directive entered into force, reservations made under it still do not reliably reach, and are not reliably detected and respected by, the systems they are addressed to. Since the AI Act requires general-purpose AI providers to put in place a policy to identify and comply with such

reservations (Article 53(1)(c)), this unreliability is no longer only a problem for rightsholders: providers must now demonstrate compliance with reservations they cannot dependably find. The underlying infrastructure problem is, however, broader. The same need for reliable, machine-readable rights information arises wherever the copyright status of works has to be established at scale: for public domain status and open licensing, and for orphan and out-of-commerce works.

The discussion about registries has moved from the margins of copyright policy to its centre.

In 2026 alone, two substantial studies have been published. In May, a study commissioned by the EUIPO mapped the existing landscape of copyright databases and metadata standards and sketched “CopyrightView,” a portal that would connect them – explicitly “not conceived⁴ as a new centralised database, nor as a replacement for existing systems, but as a federated transparency layer facilitating structured access to copyright-related information already held by national, sectoral and EU-level actors.”³

And in July, a feasibility study commissioned by DG CNECT concluded that an EU-level registry of TDM opt-outs is technically and operationally feasible, and recommended a specific architecture for it.⁴

Crucially, both studies proposed the ISCC⁵ as the content-derived identifier for this infrastructure. However, neither study resolved the question that will determine whether this infrastructure serves the European public interest: how an ecosystem of registries should be architected, and what role public institutions should play in it.

To answer this question, this paper draws upon the experience gained from the development and operation of CommonsDB over the past 18 months. CommonsDB is an EU-funded pilot project that has established [a functional public registry](#) of rights information for public domain and openly licensed works. It currently holds over 6.6 million verified declarations, making it the largest known public deployment of an ISCC-based registry to date.

The CommonsDB pilot demonstrates that purpose-built, ISCC-based registries work: technically, operationally, and at scale. What no individual registry can provide – and what neither of the 2026 studies specified – are the connections between registries that are necessary for a functioning ecosystem: a way to discover which registry knows about a given work, and a way to query a defined set of registries as if they were one.

These are the layers that require a careful institutional design that assigns public institutions distinct functions: as standards-setters and registry-certifiers and registry-aggregators, and, where the law so provides, as registry operators. New registries envisaged by policymakers and other stakeholders should not be conceptualised as central databases that absorb existing registries. Public institutions should instead certify independently operated registries as compliant for defined purposes, operate the aggregation and discovery layers that make a certified ecosystem function as a whole, and migrate its own existing copyright databases onto the same architecture. Ultimately, this is based on the insight that most registries are better built, and better governed, close to the communities that populate them.

The remainder of the paper proceeds in four steps. Section 2 sets out what the CommonsDB pilot has demonstrated, and what it has not. Section 3 identifies the two discovery problems that mark the limits of any single registry and describes the layers at which they are solved. Section 4 develops the architecture of such an ecosystem of certified registries. Section 5 draws the conclusions for the EUIPO's role in the emerging European copyright infrastructure. In doing so, it draws on the legal analysis prepared for this deliverable and [published as an annex to this paper](#). This analysis examines the actions that the EUIPO can take within its existing mandate, as well as the areas where legislative intervention is necessary.

2

What CommonsDB has demonstrated

As an EU-funded pilot project, CommonsDB aims to assess the feasibility of a public registry of works in the public domain and with open licenses. Coordinated by [Open Future](#), the project includes partners associated with two of Europe's largest repositories of openly licensed material: the [Europeana Foundation](#) and [Wikimedia Sverige](#)⁶. It also includes the [Institute for Information Law \(IViR\)](#) as research partner, and [Liccium B.V.](#) as a technical partner.

The project [was originally inspired by the automated content moderation obligations of Article 17 of the CDSM Directive](#), for which reliable public domain information is a necessary corrective. However, by the time the project began, the policy debate had moved on, and the pressing use case for machine-readable rights information had become the use of works as training data for AI models. The project adjusted its focus accordingly⁷ while the underlying purpose remained unchanged: to demonstrate that verifiable rights information can be bound to works themselves and made available as public infrastructure.

The choice of subject matter was deliberate. Public domain and openly licensed works are the one segment of the copyright landscape where every architectural question can be tested without contested rights claims, without incumbent gatekeepers, and without legal exposure for the parties involved: declarations are voluntary, the underlying data is open by definition, and errors are correctable in public. This made CommonsDB a safe space in which to test the building blocks that every currently proposed copyright registry, including a TDM opt-out registry, would rely on at production scale. That being said, nothing in the architecture is specific to these rights statements as the payload: the declaration model (identifier, signed assertion, verified identity, timestamp) can carry any assertion type, such as opt-out reservations, out-of-commerce status, or provenance claims.

At the same time, a registry of public domain and openly licensed works serves a purpose beyond that of a test bed. In the EUIPO mapping study's survey of the entire landscape, CommonsDB appears as the only initiative aimed at increasing legal certainty for the reuse of digital content. None of the existing or proposed registries – ranging from the out-of-commerce works portal to a future opt-out registry – are designed to answer the question that reusers, platforms, cultural heritage institutions and AI developers actually ask: is this work free to use? A registry of verified public domain and open licence status answers it affirmatively, and an affirmative answer moots the opt-out question for the works it covers. The positive-signal layer is the necessary complement to the emerging opt-out infrastructure and, as a result of the consortium's efforts over the past 18 months, it exists today in the form of CommonsDB.

At the end of the pilot project phase, the CommonsDB pilot has produced the following results:

- 1 A registry of verified declarations works at scale.** The CommonsDB registry holds over 6.6 million declarations submitted by the data partners, surpassing the project target of five million. Each declaration binds a rights statement to a content-derived ISCC identifier, is signed by the declaring party, and is independently timestamped. The registry stores metadata records, not works. It holds identifiers, rights statements, declarer identity, timestamps and source references, while the content itself remains where it is hosted.
- 2 Declaration pipelines are feasible for data providers.** Both data partners have built working pipelines on their own infrastructure. ISCC codes are generated locally from the media files, metadata is packaged, signed, and submitted in bulk to the declaration API. Neither partner had to surrender content, restructure their collections, or adopt a new metadata standard wholesale. This matters beyond the pilot because it demonstrates that the cost of joining a registry ecosystem lies in tooling and workflow integration rather than institutional transformation – a precondition for the distributed onboarding model that any EU-scale registry will need.
- 3 Content-based identification and matching work in practice.** Content-derived identification is the central choice in the architecture: because an ISCC is computed from the media file itself, any party holding a copy of a work can independently derive the same identifier and use it as a lookup key – no prior coordination, no central issuance, no dependence on metadata that gets stripped during circulation. The registry demonstrates

that matching against a corpus of this size works, supporting both exact matching and near-duplicate detection. Two operational lessons come with this: Firstly, matching requires conservative thresholds: content-level similarity must never be interpreted as work-level equivalence or as an implicit rights assertion – a distinction the TDM feasibility study, notably, also adopted when it dropped similarity matching from its final design. The legal analysis accompanying this paper arrives at the same rule from the law's side: because the ISCC establishes resemblance between files rather than the identity, authorship or legal status of works, a public operator that presented ISCC similarity as legal equivalence would be claiming more than the technology warrants⁸. Secondly, matching at scale requires mechanisms to reliably distinguish genuine matches from false positives⁹. The multi-component structure of ISCC supports a layered matching approach, using compact identifiers for efficient discovery and more precise content-derived information to validate and refine potential matches.

4 ISCC and verifiable credentials combine into a workable trust model.

A registry of declarations is only as trustworthy as its answer to the question: who is declaring, and how do we know? At scale, this answer must be machine-verifiable. In CommonsDB, declaring parties are identified through verifiable credentials anchored in eIDAS¹⁰-compliant certificates; declarations are signed by the declarer and countersigned by an independent time-stamping authority. The EUIPO mapping study observes that most existing copyright databases “rely on declarer trust for authenticity,” with no verification mechanism. The CommonsDB pilot demonstrates the alternative: a verifiable binding between content (through the identifier), assertion (through the signed declaration) and identity (through the credential). In section 4, we return to explain why it is this triple binding – rather than the registry mechanics – that matters most for the opt-out use case.

5 Federation works. The feasibility study committed the project to federation as a core design principle, and the pilot has delivered on its architectural promise. The registry includes a distributed layer, implemented as an open-source service over the Iroh peer-to-peer protocol¹¹, through which registry entries are published and synchronised directly between nodes. This enables any interested party holding a read ticket to synchronise the registry over the peer-to-peer network – there is no gateway and no central query service in the read path – and the entire registry has been replicated on independent nodes outside the project's

core infrastructure. Therefore, a party that needs to run lookups at scale can hold a complete, continuously synchronised local copy and resolve against it locally, at whatever volume its own hardware supports. Detailed declaration metadata remains at the source; what federates is the minimal record needed for discovery, metadata resolution, and verification. The same mechanism also works across registries: because every participating registry publishes its records in the same way, anyone can synchronise the records of several registries at once and query them as if they were a single registry. Sections 3 and 4 build on this property, making it possible to combine multiple independent registries (for example, several TDM opt-out registries) into one aggregated view.

6 Identity is the hard part. Registries require proper authentication of the declaring party: verification at scale can only be automated when every record is reliably bound to a verified identity. The CommonsDB trust model delivers this, but building it surfaced the immaturity of the infrastructure it depends on. Qualified organisational credentials remain rare among the institutions that hold relevant collections, the tooling for issuing and integrating them is limited, and acquiring and deploying credentials proved to be among the most demanding parts of the onboarding process. The lesson generalises: any EU-scale registry design must treat declarer identity as a first-order design problem rather than an implementation detail – a conclusion the TDM feasibility study implicitly shares when it dedicates its entire distributed layer to identity and mandate verification.

7 Onboarding requires incentives that a registry cannot create for itself. Beyond the committed data partners, the growth of the registry depends on declaring parties having a reason to declare. In the context of the public domain and open licensing that reason is diffuse: a declaration serves the public interest and the wider reuse ecosystem, but produces no direct effect for the declarer. In the TDM opt-out context, however, the incentive structure is reversed: a reservation record exists to produce a legal effect for the party who makes it, and Article 53(1)(c) of the AI Act gives developers equally concrete reasons to consult it. The lesson is that participation follows purpose: registries attract participants when their records benefit them in some way. This is why the certification model developed in Section 4, which makes inclusion in the certified set the reward for meeting the requirements, matters as much for adoption as for trust.

8 Registry operation is the cheap part. Usage adds variable costs for processing declarations and serving queries, in the order of €350 per million API transactions, and these are borne in proportion to demand; high-volume users can avoid them altogether by holding a complete synchronised copy of the registry, which requires no more than commodity hardware, as the pilot demonstrated. The substantial costs sit elsewhere: in developing and maintaining the software, in identity verification and onboarding, and in the human work of supporting data partners. This matters for the argument in section 4: the expensive functions are the ones that any design, centralised or federated, must place close to data providers and rightsholders, while the records layer itself is the part that is inexpensive to operate.

The pilot's most consequential insight emerged alongside the findings listed above: Over the course of the project, it became clear that the question of whether a registry could be built had been answered and that a new question had taken its place: what happens when many registries exist? We consider this a very likely scenario given the range of stakeholders affected and because the incentives, mandates and constituencies behind them differ.

CommonsDB was built and conceived as a single registry. The landscape that is emerging, and that EU policy must now design for, is an ecosystem of registries maintained and governed by different entities. The remainder of this paper addresses this overall question.

3

From single registries to an ecosystem

Both 2026 studies point towards a multi-registry world: the EUIPO mapping study explicitly states that CopyrightView should be a federated transparency layer over existing databases rather than a new central one. The TDM feasibility study conceives its registry as “a supportive and complementary instrument” that leaves source-based signals and sector identifiers in place.

The EUIPO study comes closer to naming a mechanism¹²; among the three candidate architectures it presents for CopyrightView is a fingerprint-based discovery protocol, the ISCC Foundation’s IDP¹³, which “would enable sector-specific or jurisdiction-specific registries to coexist and mutually benefit.” The study presents it as one of three candidate options for further consideration.

Fundamentally, discovery is only half of the connective tissue a registry ecosystem needs: routing an identifier to the registry that holds information about it says nothing about how a defined set of registries can be queried as one, or on whose authority the set is defined. These are two distinct problems that need to be solved at different layers and by different mechanisms and they come with different governance questions.

Two discovery problems

These two problems are worth examining in detail because they are not incidental technical difficulties: they determine how the overall system must be architected. How information is found across registries decides what has to be common to the ecosystem, what can remain distributed, and which functions need an institutional operator. The architecture this paper proposes in Sections 4 and 5 follows directly from how they are solved.

Problem 1 - discovery: I hold an ISCC code derived from a file I am about to use and I need to find out whether any registry, anywhere, holds information about it.

This is a discovery problem: with nothing but a content-derived identifier in hand, and no prior knowledge of which registries exist or which might be relevant, I need to be routed to the registry that can provide an answer. The function is precisely analogous to DNS: a neutral resolution layer that maps an identifier to the place where the corresponding information lives.

This layer is now taking shape. In July 2026, the ISCC Foundation published a set of draft specifications covering the ISCC-ID, a discovery protocol, a transparency-log design and a signature scheme, under which registries declare their holdings to shared “hubs,” whose append-only, cryptographically verifiable logs are aggregated into a cross-registry index¹⁴. A user submitting an ISCC code can find out which registry holds the corresponding declaration and where to retrieve it, without knowing that the registry exists. The design is openly specified, the reference implementations are open source, and a first deployment went live in mid-2026 – making this effort even more recent than CommonsDB.

As with any infrastructure at this stage, what exists today is a starting point rather than a finished public utility. The specifications are drafts, framed as input to a future revision of ISO 24138 or a related standard; the running components are early implementations; and the network is currently operated by those who initiated it, with real-world usage coming from a small number of early adopters. None of this is a criticism and the same can be said of CommonsDB, whose own federation layer has so far been implemented only by a handful of parties. It does, however, point to a general property of registry infrastructure: neutrality and reliability cannot be established by specification alone. They become real when independent parties operate the components, and depend on them. For a routing layer this matters with particular force, because it serves the whole ecosystem or it serves no one. This is why Section 5 argues that it is at this layer that an institutional anchor turns an open design into shared infrastructure.

Problem 2 - bulk querying against a defined set: I need to run one code (or, in the AI-pipeline case, millions of codes) against a defined set of registries.

Where the discovery problem is one code against all registries, this problem is the inverse: running against a defined set of registries. For an AI developer assembling a training corpus, the relevant set is “all registries whose opt-out records count”; for a platform resolving reuse permissions, it might be “all registries carrying verified public domain status.” The set, not the individual registry, is the unit that matters, and in many cases the query volume rules out anything that requires a network round-trip per work.

The CommonsDB federation layer has been designed to solve this problem. It does so in a way that departs from how data aggregation is traditionally understood. The dominant model of aggregation is a central service: the aggregator ingests the registries and everyone queries the aggregator. The drawbacks of this model are well known. The central service is a single point of failure; when it is unavailable, compliance checking stops across the entire ecosystem. Its capacity, availability and pricing set a ceiling on what everyone else can do, and every registry must coordinate with it, on its terms, to remain represented. Its operator acquires a gatekeeping position over whose records are visible and under what conditions. The model also breaks down at the volumes involved: for a party that needs to check millions of works, querying over the network is not a realistic option as per-work round-trips are too slow and too costly on the querying side, and serving that load is expensive for whoever operates the service.

Once registries publish their records via a synchronisable peer-to-peer layer, the aggregator’s authoritative product becomes a list rather than a query endpoint: a machine-readable, signed enumeration of the registries belonging to the set. In this design anyone holding that list can synchronise those registries and maintain a complete, continuously updated local replica of the aggregated view. The records themselves remain with the registries that produced them, keeping the overall architecture in line with the principle that data should be stored as close as possible to the party from which it originates. At the same time, each user can hold as much of the aggregated view as their own needs justify.

The result is two interaction models that serve different types of users, with different needs and resources, from the same data. Occasional users, such as a rightsholder checking the status of a work, a platform handling a dispute, a court verifying a record, query a public search interface, maintained and provided as public digital infrastructure.

High-volume users, such as AI developers running millions of lookups inside their pipelines, synchronise the relevant set of registries and resolve identifiers locally, at whatever speed their own infrastructure supports. Both get identical answers, because both operate on the same synchronised records. The search interface is simply one instance of a service that anyone holding the list can run.

4

A federated architecture for copyright-information registries

As noted in the introduction, EU copyright law has, following Berne, traditionally been reluctant to rely on registries. Registry-like structures have nonetheless begun to appear at the edges of the *acquis*: the Orphan Works Directive introduced the orphan works database, and the CDSM Directive introduced the out-of-commerce works portal, both operated by the EUIPO. With the challenges raised by generative AI, this cautious opening has turned into active interest in registries of rights information more generally.

Interest in registry-based approaches is building across the institutional landscape. The EUIPO has looked at them twice: its 2025 study on generative AI and copyright surveyed emerging registry solutions in detail, including the federated architecture of sector-specific, synchronised registries on which CommonsDB is built, and its 2026 mapping study, discussed above, makes registry infrastructure a central element of the CopyrightView proposal. The first commercially operated registries, built on the same core building blocks that CommonsDB has validated at scale, are already on the market.

For the architecture of such an ecosystem, two basic models are available. In a centralised model, declarations flow from distributed onboarding points into a single canonical index; every query passes through one read API; and one operator issues the responses and receipts on which the whole ecosystem relies. In a federated model, the records remain in purpose-certified registries, governed by the constituencies that populate them; a signed, machine-readable list maintained by a public certifier defines which registries count; the records synchronise over a common peer-to-peer protocol; and the public

search interface is one replicable instance of a service that anyone holding the list can run. This paper argues for the federated model, for reasons of governance rather than engineering.

Before making the case, it is worth setting out the two architectures side by side, together with the small number of concepts on which the rest of this section relies.

Two approaches to building a European registry infrastructure:

The centralised model. Declaring parties submit their declarations through sectoral or national onboarding bodies, which verify their identity and their mandate. All verified declarations then flow into a single central registry, operated by one institution. Everyone who wants to check what is on record for a work – an AI developer, a platform, a court – queries this central service, which provides the authoritative answer and a receipt confirming that the check took place.

The federated model (this paper's proposal). Declaring parties submit their declarations to any of several certified registries – run by sector bodies, member states, collective management organisations or service providers – each of which verifies identity and mandate to the same common standard. The registries continuously share their records with each other and with anyone else who wants them, over a common synchronisation protocol. The role of the public institution operating the system is not to hold the records; it is to certify the registries and to maintain the authoritative list of those that qualify. In addition it could operate a public search interface across all of them.

The concepts used below:

Registry certification – the process by which a public authority confirms that a registry meets the common requirements (identity verification, data model, record-keeping), so that its records count for compliance purposes. A registry that passes is a **certified registry**; together they form the **List of certified registries**.

List of certified registries – the machine-readable, signed list of certified registries, maintained by the certifying authority. Whoever holds the list knows exactly which registries to consult. There can be multiple lists serving different use cases.

Synchronisation and local copies – because certified registries publish their records over a common protocol, any sufficiently sophisticated party can maintain a complete, continuously updated

copy of all records in the certified set on its own systems, and check works against it at any speed their infrastructure allows.

Signed checkpoints – cryptographically signed statements of what the certified set contained at a given moment, published at regular intervals. Each certified registry signs its own records; the public institution that maintains the certified list publishes the checkpoints covering the certified set as a whole. Checkpoints let anyone prove, after the fact, what was (and was not) on record at a particular date.

The discovery layer – the routing function described in Section 3: a way to find out, starting from nothing but a content-derived identifier, which registry holds information about a work.

Records belong close to those who make them

The starting point of our argument in favour of the federated model is the logic of proximity: identity checks and declaration workflows belong with sectoral and national bodies because that is where the knowledge and the relationships are: a music-sector body understands music-sector mandates. But the same is true of the records themselves.

A registry for the music sector will make different choices about workflows, metadata and granularity than one for scientific publishing or for visual works, and the body that can verify a declarer's mandate is also the body best placed to govern the records that declarer produces, to correct errors, and to answer for the registry's contents to the community it serves.

A single central registry does not make these differences go away. It flattens them into a ¹⁵lowest common denominator, and moves every contested question (whose record wins, what counts as a valid mandate, how mistakes are corrected) into one institution that is distant from all of the affected constituencies at once. Keeping records with those who produce them avoids both problems. What individual registry operators cannot supply for themselves is mutual discoverability, common standards, and an authority that says which registries count. This, rather than record-keeping, is the public function.

The federated model can deliver what registry users need

The case for a central registry rests on what the users of the system need: one clear answer per work, evidence they can rely on, and knowing how current that answer is. The federated model delivers all three.

The answer is the same everywhere, because the data is the same everywhere. Whether a developer queries the public search interface or checks works against a synchronised local copy, both draw on the identical records of the identical certified set.

How current an answer is follows from the signed checkpoints, which the operator of the aggregation layer publishes at regular intervals across the combined records of the certified set, and which state exactly which state of the records an answer reflects. The checkpoints also supply the evidence: anyone who relies on the records can retain them and later demonstrate what was, and was not, on record at a given date, and can do so verifiably, because checkpoints can be checked independently against the public record rather than having to be taken on trust.

Should two certified registries ever disagree about the same work, the rules for resolving the conflict – which record takes precedence, and a cautious default while a dispute is pending – are set centrally by the certifying authority, exactly as they would be in a central registry. Nothing about those rules requires the records themselves to be held centrally.

Certification: what it takes for a record to count

There is a deeper reason to put public institutional energy into certification rather than record-keeping, and it comes from asking what actually happens when someone relies on a registry record, whether in a dispute, an audit or a reuse decision. Two facts matter: who made the declaration, and what it said at the relevant point in time (for example, on the date on which a public domain image was included in a publication). A registry entry that cannot prove these two facts may be technically sophisticated and still be of little use when it matters most.

Securing them takes more than cryptography. It requires three things, and they define the floor that certification must set for any registry whose records are to be meaningful.

- 1** First, **verified identity**: the declarer's identity must be established against an eIDAS-anchored credential, not merely asserted, so that a record proves who declared rather than that someone claiming a name did¹⁶.
- 2** Second, **self-contained records**: the substance of the declaration must be machine-readable and bound into the signed, logged record itself – not referenced through links to documents that can be changed later – so that what the declaration said at any moment is beyond argument.
- 3** Third, **versioned histories**: records must be kept in append-only, versioned form, so that the state of a record on any past date can be proven, including, for example, by a rightsholder who wants to show that a reservation was in place before a given AI training run.

Meeting these **three requirements** is within reach of any well-run registry – one of the key achievements of the CommonsDB pilot has been to demonstrate all three at scale, so the requirements are demonstrably practical rather than aspirational.

What no registry can supply for itself is the legal significance of its records: whether a developer, a regulator or a court may treat them as reliable evidence of a valid declaration. Only a public authority can confer that significance, and certification is the instrument for conferring it. It turns the three requirements from good practice into conditions of participation, for public and private registries alike.

The envisaged architecture

Concretely, the registry infrastructure we propose looks as follows. We consider the EUIPO the possible host for the public functions this architecture requires.

Records live in multiple registries (sectoral, national, collectively managed or commercially operated) each governed by and answerable to the constituency whose reservations it records. Such registries do not need to be created by decree: the first ones have already emerged on their own, and more will follow as the legal and practical weight of registry records grows. What must be put in place deliberately is the certification framework around them. The natural sequence is that a registry gets established and populated first, and certified once it meets the requirements. This is why the criteria should be available as early as possible: the sooner operators know what will be required of them, the more registries will be designed to meet the requirements from the start.

The EUIPO could define those criteria: the identity and mandate verification rules; a data model that is common, FAIR and versioned, so that it can start minimal and be extended as richer vocabularies for specific purposes and sectors mature; the evidence floor described above; and publication of records over the common synchronisation protocol. The EUIPO certifies registries against these requirements, maintains the certified list, and operates a public search interface, in the knowledge that the interface is one instance of a replicable service and that any party can synchronise the certified set and resolve locally.

Two extensions complete the picture:

- 1** The same search interface can federate different types of sources: the [out-of-commerce works portal](#) the EUIPO already operates, verified positive-signal registries such as [CommonsDB](#), whose affirmative answer – this work is public domain – settles questions about its reuse, and, where they exist, registries of rights reservations.
- 2** And the certified registries publish their holdings in the discovery layer, which is the routing function that the ISCC Discovery Protocol introduced in Section 3, so that anyone holding nothing but an ISCC code can find their way to the right registry without knowing in advance that it exists.

No element of this architecture needs to be invented. The identifier is an ISO standard; the trust model runs on eIDAS; the synchronisation layer is demonstrated at scale in CommonsDB; the discovery layer is specified and in early deployment; and the first candidate registries already exist. What is missing is the connective decision: the certification regime and the public anchors that turn these components into one infrastructure.

Costs and incentives

The public layers of this architecture are not free. Maintaining the certification process, the certified list, the synchronisation infrastructure, the public search interface and a discovery-layer hub involves real, ongoing costs. But these are thin layers compared to operating a central registry for all of Europe's reservations, and ¹⁷it is reasonable to expect their cost to be lower than – or at worst in line with available estimates for operating a fully centralised registry.

On the other hand, the operation of the records layer itself no longer falls on the public budget. That layer is carried by registry operators – and the reason to expect them to carry it is that the incentives are real and, in several cases, already acting. Collective management organisations have every reason to operate reservation registries on behalf of their members, for whom expressing and defending reservations is core business. Publishers and picture agencies have catalogs to protect and existing rights-management infrastructure to build on. Service providers have commercial reasons: two ISCC-based opt-out registries are already on the market, ahead of any EU decision. On the positive-signal side, cultural heritage institutions and open-content communities have demonstrated their willingness to declare at scale through CommonsDB. The common European data space for cultural heritage would benefit from a verifiable rights-information layer. Certification of registries by an authoritative public institution would strengthen all of these incentives at once, because it attaches the one thing none of these actors can create for themselves: official recognition that their records count. Where records carry legal or practical weight, as they do for rights reservations under Article 4(3) and the AI Act, records that count are worth maintaining. The reverse holds as well: records held in registries that do not meet the requirements do not count, however carefully they were made.

One caution belongs in any EU-institutional endorsement of this architecture: append-only public records, in which nothing is ever truly deleted, sit uneasily with data protection law and with the legitimate interest of rightsholders in changing their minds. This question lies outside the scope of [the legal analysis](#)

[accompanying this paper](#), which notes the data protection obligations that apply to an EU-institutional operator but does not examine them in depth. The interaction between permanent public logs, the GDPR and the revocation of reservations therefore remains an open question, and it deserves dedicated analysis before any EU body adopts append-only records for registry purposes.

5

The EUIPO's possible role: certify, aggregate, anchor – and migrate

The EUIPO is the natural institutional home for the public layers of this architecture. It already operates the EU's two existing copyright registries – the out-of-commerce works portal and the orphan works database – under express legislative mandates.¹⁸ Through the [Copyright Knowledge Centre](#) it has built the institutional apparatus for copyright infrastructure work, and its CopyrightView proposal points in the right direction: a federated transparency layer rather than a new central database. The European Parliament has reached a similar conclusion from the policy side: its 2026 resolution on copyright and generative AI proposes an intermediary role for the EUIPO in managing and listing rights-reservation mechanisms that can refer to catalogues already in existence. What is still missing from both is an architecture. This paper's proposal is, in substance, that architecture under which the EUIPO would be assigned four tasks:

- 1 Registry-certifier.** The EUIPO would define and apply the certification criteria for purpose-compliant registries and maintains the authoritative, signed, machine-readable lists of certified registries per purpose. The criteria should combine the evidence floor of section 4 with a structural test that we propose as the general standard for every component of this infrastructure: openly specified, independently operable, and governed outside the control of any single operator. The legal analysis accompanying this paper arrives at the same requirement from a different direction: public adoption of infrastructure that exists only as a single-vendor implementation risks dependence on proprietary systems and weakens the public-law legitimacy of the arrangement.¹⁹ Because registries will

emerge and mature organically, the certification criteria should be published early – well before most of the registries they will apply to exist – so that new registries can be designed against them from the start.

- 2 Registry-aggregator.** The EUIPO would operate per-purpose search interfaces that can be used by rightsholders, developers or courts to check any work against the certified set.²⁰ It also publishes the signed checkpoints that state, at regular intervals, what the certified set contains. It does so in the knowledge, and with the stated intention, that the interface is replicable: the authoritative artifacts are the certified list and the synchronised records, and any party may run the same service against them. The EUIPO's interface would derive its status from the certification behind it rather than from exclusivity. Taken together, certification and aggregation produce the outcome that the whole architecture exists to deliver: one authoritative, verifiable view over many independently governed registries.
- 3 Anchor of the discovery layer.** The discovery layer that resolves an ISCC code to the registry that holds information about it must be credibly neutral and predictably permanent if it is to serve the whole ecosystem. As set out in section 3, the ISCC Discovery Protocol provides a credible, openly specified design for this layer, but it is still maturing, and its operation has not yet been institutionalised. The EUIPO should anchor it: operate a hub, participate in the protocol's governance so that it meets the same structural test, and thereby resolve the chicken-and-egg problem that a discovery layer becomes useful only once many registries publish into it. An institution that certifies registries is the natural bootstrap for their mutual discoverability – which is also why anchoring the discovery layer belongs alongside certification, rather than ahead of it.
- 4 Migration of its own registries.** Finally, the EUIPO should bring its existing databases into the ecosystem it anchors: the out-of-commerce works portal and the orphan works database should, over time, be re-founded as ISCC-based registries of verified declarations, publishing over the same synchronisation and discovery layers and meeting the same certification floor. Here the EUIPO legitimately runs registries: where a legal mandate designates it as their home. The general principle stands: the EUIPO operates registries where Union law makes it the registrar, and certifies, aggregates and anchors everywhere else.

The same analysis points **to three risks** that the institutional design should avoid:

- 1 Centralisation by default.** Responding to new registry needs by building new central databases would reproduce, at the EU level, the governance problems described in the preceding section: records concentrated in an institution distant from the communities that produce them, and sectoral difference flattened into a lowest common denominator.
- 2 Absorption of community registries.** Folding existing registries into the EUIPO's own operation would extinguish what makes them valuable. The value of a registry lies in its proximity to its community, and absorption severs exactly that proximity.
- 3 Lock-in through procurement.** Treating copyright infrastructure as a procurement problem – a system specified once and contracted out – risks tying the public function to a single implementation and a single vendor. What is being built is an ecosystem whose components will be operated by many hands, and the durable public contribution consists of the standards, the certification and the anchors rather than any single system.

The legal analysis reinforces this allocation: under the principles governing what EU agencies may do, the stronger the legal and practical effects of a function, the stronger the need for an express legislative mandate²¹. Operating the authoritative record of Europe's rights management registries sits at the heavy end of that spectrum; "certifying, listing, aggregating and anchoring" sit progressively at lighter ends of the spectrum. The architecture proposed here is, in other words, the one most compatible with the EUIPO's existing legal position.

This also answers the question of how to begin. According to the legal analysis, the EUIPO can already, under existing law, support, pilot and even host voluntary, non-constitutive copyright-information services – provided they remain declarative, transparent and free of legal effects on protection – especially where this happens in cooperation with the Commission and within its approved work programme²². The certification criteria, the certified list, the reference interface and a discovery hub can all be developed on this basis. A durable regime will require legislation, and the legal analysis sketches the available paths: amending Regulation 386/2012, amending the CDSM Directive, or a targeted instrument on copyright-information infrastructure that would place the certified registry

ecosystem alongside the orphan works and out-of-commerce works mandates. Nothing about that legislative horizon prevents the work from starting now, as an explicit development and pilot phase whose status is clear to everyone taking part. What the horizon defines is where the work should arrive.

The legal analysis also defines the safeguards under which this can happen: the services must remain declarative, voluntary, non-constitutive, technically transparent, open to audit and limited in their legal claims. It further specifies what a future legislative mandate would need to define: the purpose of each registry and the extent of legal effects of its records, admission and suspension criteria and the status of the certified list, conflict-reporting and correction procedures, data protection responsibilities, liability allocation, interoperability requirements and funding, with the clarification that none of this introduces copyright formalities or affects the availability of works outside the registries. And it recommends that, whichever legislative path is chosen, the ISCC identifier layer be recognised as a common technical foundation and the registry and discovery layers be developed as open specifications, with the EUIPO operating a reference implementation and enabling member states, cultural heritage institutions, data spaces and other public infrastructure providers to run interoperable nodes.

6

Conclusions

The CommonsDB pilot set out to test whether a public registry of rights information including content-derived identifiers, verified declarations, and credential-based trust could be built and operated at scale. We have shown that it can: more than 6.6 million verified declarations, working declaration pipelines at two of Europe's largest repositories of openly licensed material, demonstrated federation across independently operated nodes, and modest running costs confirmed the technical feasibility.

The two studies on copyright infrastructure published in 2026 confirm that the institutions have reached the same conclusion on the fundamentals: content-derived identification, verified declarations and registry infrastructure are the direction of European copyright infrastructure.

The question that remains open is architectural, and it will be decided either by deliberate choice or through the design of the first systems that get built. A centralised records layer would decide it one way. This paper has argued for the other: records in purpose-certified registries governed close to their constituencies; discovery, aggregation and certification as the public layers; the EUIPO as certifier, aggregator and anchor rather than as registrar of first resort. The EUIPO would run registries only where the law makes them its own, and migrate those onto the common architecture.

For a strategy paper written from inside a registry project, our central recommendation reaches beyond any single registry: the European Union should build the layer that only it can build; the certification regime, the authoritative lists, and the neutral anchors that turn independent registries into shared copyright infrastructure. The building blocks exist: a content-derived identifier that is an ISO standard; a trust model anchored in eIDAS; a federation layer demonstrated at scale; a specified discovery protocol; the first operating

registries; and an institution whose existing mandates and competences can carry the first phase of the task. What is at stake also reaches beyond copyright administration. An openly specified, publicly anchored infrastructure for rights information is a piece of European digital sovereignty: machine-readable knowledge about European works, held as open public infrastructure rather than accumulating exclusively in the data portfolios of the largest platform and AI companies²³. The time to assemble them into a coherent and sufficiently flexible European copyright infrastructure framework is now, before the design of the first EU-level systems fixes decisions that will be expensive to revisit.

Notes

¹ Article 5(2) of the Berne [Convention for the Protection of Literary and Artistic Works](#) (Paris Act of 24 July 1971, as amended on 28 September 1979): “The enjoyment and the exercise of these rights shall not be subject to any formality.” The prohibition concerns conditions placed on the enjoyment and exercise of copyright; it does not preclude voluntary, non-constitutive registries of rights information, in which registration carries no consequences for protection – see the legal annex to this paper, section 3.3.

² The Copyright Infrastructure Task Force is hosted by the Finnish Ministry of Education and Culture. See its First Project Report (OKM 2025:23) and the 2026 status-and-way-forward report.

³ EUIPO, [Mapping Study: EU Databases and Metadata Standards for Copyright Information](#), May 2026.

⁴ Visionary Analytics / Capgemini Invent for the European Commission (DG CNECT), *Study to Assess the Feasibility of a Registry of Text and Data Mining Opt-outs Expressed by Rightsholders*, July 2026.

⁵ ISO 24138:2024, Information and documentation – International Standard Content Code (ISCC), published May 2024. The standard was developed on the initiative of the ISCC Foundation; open-source reference implementations are available. The ISCC’s role in the CommonsDB architecture is described in section 2.

⁶ Technically the Wikimedia Commons is operated by the Wikimedia Foundation, in the context of CommonsDB, Wikimedia Sverige is operating as a bridge to the community that maintains the Wikimedia Commons.

⁷ The upcoming final report will give a fuller account of this shift.

⁸ Legal annex, section 3.3 (proportionate technical operation as a limit on public operation of ISCC-based infrastructure).

⁹ As the CommonsDB registry has grown, the 64-bit content codes increasingly produced false positives, and more robust matching at ecosystem scale will require the full 256-bit ISCC units.

¹⁰ Regulation (EU) No 910/2014 on electronic identification and trust services for electronic transactions in the internal market (eIDAS), as amended by Regulation (EU) 2024/1183 (eIDAS 2.0). The eIDAS framework provides legally recognised means of establishing and verifying the identity of natural and

legal persons across the EU, including qualified certificates and electronic seals for organisations. In CommonsDB, these anchor the verifiable credentials with which declaring parties sign their declarations: the credential proves which organisation made a declaration, with the same legal standing across all member states.

¹¹ CommonsDB iroh registry service, <https://github.com/CommonsDB/commonsdb-iroh-service> (Apache-2.0/MIT)

¹² EUIPO study §4.4.3, pp. 114–115; the “presented for further consideration” language is on p. 108 and again in the conclusion.

¹³ ISCC Foundation, ISCC Enhancement Proposals: IEP-0011 (ISCC-ID), IEP-0013 (ISCC Discovery Protocol), IEP-0014 (Transparency Log) and IEP-0019 (ISCC Signature), <https://ieps.iscc.codes> (all Draft status; last updated July 2026; accessed August 2026). The specifications are framed as input to a future revision of ISO 24138.

¹⁴ ISCC Enhancement Proposals IEP-0011 (ISCC-ID), IEP-0013 (Discovery Protocol), IEP-0014 (Transparency Log) and IEP-0019 (ISCC Signature), <https://ieps.iscc.codes>.

¹⁵ The final report considers and rejects more expressive reservation models, on the ground that purpose-based permissions and distinctions between AI system types “would introduce ambiguity and make automated enforcement impractical in developer pipelines” (TDM feasibility study, p. 84). The interim design, presented at the study’s multi-stakeholder validation workshop in December 2025 – in which members of the CommonsDB consortium participated – still provided for reservations differentiated by purpose, territory and duration. The move to a strictly binary record was made between the interim design and the final report.

¹⁶ Parts of the system can also be designed to be privacy preserving. The federated model also supports pseudonymous declarations, where the local declaration service would know the identity behind the public key used to sign the declaration but does not publish it.

¹⁷ TDM feasibility study, cost estimates and funding implications (pp. 111–112): €4–8 million for a two-year pilot, €2–4 million per year steady-state central operation, €10–16 million over five years.

¹⁸ On the two mandates – Article 3(6) of the Orphan Works Directive and Article 10 of the CDSM Directive, both implemented through Regulation 386/2012 – see the legal annex to this paper, section 2.

¹⁹ See Legal annex, section 4.2

²⁰ Technically such cross registry search interfaces can also be operated by other entities on the basis of the same data sources.

²¹ Legal annex, section 3.2, discussing the Meroni line of case law and its application to EU agencies.

²² Legal annex, section 5

²³ See the legal annex, section 3.2, drawing on M.R.F. Senftleben, "Improving music metadata: towards a central repository of creative industry, online platform and AI data resources," *Journal of Intellectual Property Law and Practice* 21 (2026), 384